

- 7 Саликжанова Ш.Б., Сарекенова К.К.. Символическое значение женской одежды и украшений // Вестн. КУ. Сер.фил. – 2023. – № 1. – С. 53-63.
- 8 Wierzbicka, A. (2025). Semantics and Religion: Christian Themes, the nsm Approach. *Cognitive Semantics*, 10(3), 289-295. <https://doi.org/10.1163/23526416-bja10071>
- 9 Alefirenko N.F., Chumak-Zhun I.I. Axiological connotations of Slavic ethnonyms. *Rusin*. 2023. № 72. pp. 178–195. DOI: 10.17223/18572685/72/8
- 10 Goetz D., Haensch G., Wellmann H. Langenscheidts Großwörterbuch. Deutsch als Fremdsprache. Berlin; München, 1997. – 1182 p.
- 11 Wahrig G. Deutsches Wörterbuch von Renate Wahrig-Burfeind. München, 2011. – 1730 s.
- 12 Қазақ тілінің түсіндірме сөздігі. Ред. Т. Жанұзақов. Алматы: Дайк-Пресс, 2008. – 968 б.
- 13 Sabitowa M. Nationalkulturelle Elemente in der Semantik der Phraseologismen des Deutschen und Kasachischen // Das Wort. Moskau – Berlin, 1987, S. 296-300.

Материал поступил в редакцию журнала 17.02.2025

MFTAP 16.31.21

DOI: [10.59102/kufil/2025/iss3pp94-108](https://doi.org/10.59102/kufil/2025/iss3pp94-108)

Қ.Б. Слямбеков¹, А.А. Солтанбекова¹, Б.С.Шалабай², Т. Қалибекұлы³

¹А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, 050040, Қазақстан

²Ш.Уәлиханов атындағы Көкшетау университеті, Көкшетау, 020000, Қазақстан,

³Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университеті, Алматы, 050022, Қазақстан

ҚАЗАҚ ТІЛІНДЕГІ КҮРДЕЛІ ЕТІСТІКТЕРДІ ЖАСАНДЫ ИНТЕЛЛЕКТІГЕ ҮЙРЕТУ ТӘЖІРИБЕСІ: ТЕГТЕУ МЕН МОРФОЛОГИЯЛЫҚ МОДЕЛЬДЕУ ҮЛГІЛЕРІ

Мақалада қазақ тіліндегі күрделі етістіктерді автоматты тану мен корпуста белгіленім жұмыстарын жүргізу арқылы генеративті ЖИ модельдерін тұрғызуға арналған аннотациялау, тегтеу кезеңдерін жүзеге асырудың тәжірибесі баяндалған. Құранды етістіктер, құрама етістіктер, аналитикалық формалы етістіктерді тегтеудегі қиындықтар, ғылыми мәселелер, оларды шешудің жолдары, сондай-ақ белгіленім жұмыстарының ерекшеліктері жайлы айтылады. Үлкен тіл модельдерінің архитектурасында сол тілдің грамматикасы, ал оның трансформациясында сол тілдің грамматикалық модельдері жататыны сөзсіз. Заманауи лингвистиканың даму бағыты да осындай модельдерді әзірлеп, оны жетілдірумен байланысты болмақ. Сондықтан қазақ тіліндегі күрделі грамматикалық компоненттердің бірі – күрделі құрамды етістіктерді модельдеу, оның морфологиялық белгіленім үлгісін ұсыну – қазақ тілін автоматтандыру жұмыстарындағы ең өзекті мәселелердің бірі болып отыр. Аталмыш мақала осы бағытта жүзеге асырылып жатқан тәжірибелерді бөлісу мақсатында жазылып отыр.

Кілт сөздер: етістік, күрделі етістіктер, құранды етістік, аналитикалық формант, құрама етістік, тұрақты күрделі етістік, жасанды интеллекті, автоматтандыру, белгіленім

НЕГІЗГІ ЕРЕЖЕЛЕР

Генеративті жасанды интеллектінің Үлкен тіл модельдерін (LLM – Large Language Model) әзірлеудің трансформациясында сол тілдің күрделі грамматикалық жүйесі жататыны, ал мұндай ауқымды машинаны үйрету жұмыстары жиналған мол деректерді (big data) аннотациялау жұмыстары арқылы оқытылатыны белгілі.

Бұл зерттеу жұмысы қазақ тілінің күрделі грамматикалық құрылымдарының бірі – күрделі құрамды етістіктерді автоматты тану мен оны морфологиялық анализаторға арнап аннотациялау әдістеріне негізделеді. Мақалада қазақ тілінің күрделі морфологиялық жүйесін үлкен тіл модельдеріне үйретудің жолдары және аннотациялау тәжірибесі қарастырылады.

Зерттеу барысында күрделі құрамды етістіктердің түрін саралауға академик Қ.Ысқақовтың бөлінісі негізге алынып, бес түрі ажыратылды: құранды етістік, құрама етістік, аналитикалық формалы етістік, тұрақты етістік және салт-дәстүр етістіктері. Зерттеудің теориялық негізі ретінде А.Ысқақов, Ы.Маманов, С.Исаев, Н.Оралбаева, Б.Қасым сынды қазақ тіл білімінің көрнекті өкілдерінің еңбектері пайдаланылды. Сондай-ақ, шетелдік тәжірибе үлгісі ретінде ағылшын тіліндегі Oxford Phrasal Verbs Dictionary-дің грамматикалық және семантикалық құрылымы салыстырмалы түрде қаралды. Бұдан бөлек, агглюнативті тілдерді морфологиялық тегтеу және лемматизациялау бойынша шетелдік зерттеулер, атап айтқанда, О.Тороков пен R.Agerri (2024) зерттеуі қолданылды.

Мақаланы жазу барысында алынған мәліметтер болашақта автоматты синтаксистік тану жүйелерін жетілдіруге, Қазақ тілінің Ұлттық корпусының семантикалық және морфологиялық белгіленімін жетілдіруге, қазақ тіліндегі күрделі етістіктердің жүйелі тегтелген базасын құруға әрі сөздігін жасауға бағытталады.

КІРІСПЕ

Қазіргі дәуірде жасанды интеллектінің дамуы ғажайып деңгейге жетіп, түрлі қоғам және ғылым салаларына кіріктіріле бастады. Ал жасанды интеллектінің негізінде адам тілін өңдеу мен түсінуге бағытталған грамматика, семантика және прагматика, тілдік логика сынды жетілген тілдік технологиялар жатыр. Алайда қазақ тілі секілді агглюнативті тілдерді ЖИ-ге үйрету мәселесі өзекті болып қала береді. Сол себепті қазақ тіліндегі күрделі етістіктерді автоматтандыру – жасанды интеллектіні жасау үшін маңызды әрі көпқырлы міндеттердің бірі. Бұл тақырыпты таңдауымыздың басты себебі – қазақ тіліндегі күрделі етістіктер құрылымдық және семантикалық, функционалдық жағынан өте күрделі компоненттердің бірі. Оны автоматты тану қиын. Мәселен, Қазақ тілінің Ұлттық корпусы бірқатар күрделі етістіктерді жүйе күрделі сөз емес, дара сөз ретінде таниды. Мысалы, корпустан *сал* көмекші етістігін теріп, шыққан нәтижелерге көз жүгіртсек, корпустың күрделі етістіктерді ажырата алмайтынын байқауға болады. Екі сөздің бірігіп, бір мағынаны беріп тұрғанын түсініп тұрған жоқ. Атап айтқанда, *көре салу*, *көз салу*, *ақыл салу*, *айқай салу*, *қырып салу* сынды күрделі етістіктердің құрамындағы «сал» компоненті жеке сөз ретінде белгіленіп тұр. Осы секілді олқылықтар сөздің, сөйлемнің, мәтіннің мағынасын қате түсінуге әсер етуі мүмкін.

Сондықтан корпуспен айналысатын тілтанушылар үшін күрделі сөздерді, күрделі етістіктерді автоматты тану алгоритмдерін, жолдарын табу және жүзеге асыру кезек күттірмейтін мәселе болып отыр. Оның үстіне, агглюнативті тілдер морфологиялық құрамы біршама күрделі тілдерге жатады, сондықтан есептеуші технологияларға тегтелген грамматикалық белгілерді үйрету арқылы көптеген нейронды оқыту жүйелерін құруға мүмкіндік береді [1, 172].

Ал морфологиялық белгіленімдері кеңейтілген, семантикалық жақтан бейімделген корпус жасанды интеллектіні жасақтауға алып база болмақ. Морфологиялық мәліметтері мен леммасы енгізілген, аннотацияланған үлкен корпустардың пайда болуынан генеративті ЖИ-дің үлкен тіл модельдерінің архитектурасы құралған [1, 161].

Сол себепті зерттеудің өзектілігі Қазақ тілінің Ұлттық корпусын жетілдіріп, қазақ тілінің грамматикасына бейімделген үлкен тіл модельдерін жасау қажеттігінен туындап отыр. Қазіргі уақытта қазақ тілінің ұлттық корпусында күрделі етістікті автоматты тану бағытындағы жұмыстар біршама қолға алына бастады.

Бұл зерттеудің мақсаты – күрделі етістіктерді морфологиялық модельдеудің тәжірибесімен бөлісіп, генеративті ЖИ-ді үйретудің кейбір тәсілдерін ұсыну.

Зерттеу нысаны – қазақ тіліндегі күрделі құрамды етістіктің түрлері мен құрылымдары.

Зерттеудің гипотезасы – машинаны үйрету жұмыстарын жеткілікті деңгейде жүзеге асыру үшін қазақ тіліндегі күрделі етістіктердің морфологиялық модельдерін толықтай аннотациялау қажет деген пікірдеміз.

МАТЕРИАЛДАР МЕН ӘДІСТЕР

Мәтінді автоматты танудың әлемдік тәжірибесінде әрдайым тілдік база, оның ішінде сөздіктер маңызды рөл атқарады. Сол себепті белгі-кодтардағы ақпаратты пайдалана отырып, қазақ тіліндегі тіркескен күрделі етістіктер сөздігін жасау да – осы жобаның аясында жүзеге асуы қажет маңызды істердің бірі. Күрделі етістіктер сөздігін әзірлеудің үздік үлгісі – Oxford Phrasal Verbs Dictionary for learners of English (Ағылшын тілін үйренушілерге арналған Оксфорд фразалық етістіктер сөздігі) [2]. Бұл сөздіктің басты ерекшелігі – күрделі етістіктердің жай ғана тізімін беріп қоймай, әр тілдік бірлік туралы ауқымды әрі маңызды лингвистикалық ақпаратты қамтуы. Мәселен, сөздікте мынандай лингвистикалық мәліметтер берілген.

Негізгі мәліметтер:

- *кілт сөз* (күрделі сөзге негіз болған алғашқы компонент, дара етістік);
- *күрделі етістік* (кілт сөзден тараған барлық күрделі етістік бір ұяда беріледі);
- *күрделі етістіктің дефинициясы*;
- *мысалы*;
- *грамматикалық паттерні* (күрделі етістікті құраушы компоненттер туралы грамматикалық ақпарат, жасалу жолы).

Қосымша мәліметтер:

- *синонимдері немесе антонимдері көрсетіледі*;
- *жиі қолданылатын сөзалды бөліктері (препозиции) туралы ақпарат беріледі*;
- *аталған етістік жиі қолданылатын фразалар беріледі*;
- *аталған етістікке қатысты идиомалар көрсетіледі*;
- *қандай жағдайда қолданылатыны туралы грамматикалық ақпарат беріледі*.

Оксфорд сөздігінде күрделі етістіктің бес түрі қамтылған:

A) *idiomatic verb + particle combinations (put up with somebody)* – *идиомалық етістік + қосымша компоненттердің тіркесімдері*;

B) *verbs which are followed by a particle in a particular meaning (rely on)* – *белгілі бір мағынадағы бөлшекпен тіркесіп келетін етістіктер*;

C) *verbs that are followed by a particle in a particular meaning (brush sth up)* – *белгілі бір мағынадағы бөлшекпен тіркесетін және арасына сөз ене алатын етістіктер*;

D) *verbs with a particle, where the particle adds to, but does not change the main meaning of the verb (finish off)* – *қосымша тағы бір тұрақты компонентпен тіркесетін, алайда негізгі мағынасы дамығанымен, толық өзгермейтін етістіктер*;

E) *verbs plus particles where each has their normal meaning (phone back)* – *екі сөздің мағынасын сақтай отырып тіркесетін етістіктер* [2, 3].

Сөздіктегі күрделі етістіктердің жоғарыдағыдай классификациясы және толық аннотациясының болуы ағылшын тіліндегі күрделі етістіктерді автоматты тануды жеңілдеткені сөзсіз.

Күрделі етістік бірнеше етістікті біріктіретіндіктен мағынасы ғана емес, құрамы да күрделі және бұл күрделілік сипаты қазақ тіліндегі етістіктерді автоматты тану процесінде де айтарлықтай қиындық туғызады. Соның ішінде тіркескен күрделі етістіктердің мәселесіне жеке назар аудару қажет. Қосарлану және бірігу тәсілімен жасалған етістіктер қазақ тілінің түсіндірме сөздігінде қамтылғандықтан, біртұтас лексема және бір токен ретінде танылады. Алайда қазақ тіліндегі тіркескен етістіктерді толық жинақтаған арнайы сөздіктер жоқ. Сол себепті күрделі етістіктің бұл түрін тізімдеу үлкен жұмыс күшін талап етеді. Сондықтан оны біртіндеп жүзеге асыруды көздеп отырмыз. Агглюнативті тілдер автоматтандыруға, тегтеуге біршама ыңғайлы келеді. Өйткені қазақ тілінде түбір мен қосымшаның жігі анық көрініп

тұрады [3, 191]. Етістіктер формасы жағынан табиғатынан тұрақты не тұрақсыз болып келеді, тұрақты формалардың қатарына осы күрделі етістіктердің түрлері жатады [4, 123].

Зерттеу жұмысы барысында жинақтау, лемматизация, контекстік талдау және морфологиялық және синтаксистік модельдеу әдістері қолданылды.

НӘТИЖЕЛЕР

Қазақ грамматикасындағы күрделі құрамды етістіктердің кең тараған классификациясы – ғалым А.Ысқақовтың жіктеген топтары. Атап айтар болсақ, *құранды күрделі етістік, құрама күрделі етістік, тұрақты күрделі етістік және сараламалы күрделі етістік*. Ғалым күрделі етістіктердің түрлеріне былайша сипаттама береді:

1. Құранды күрделі етістік. Бірінші компоненті есім негізді сөз (негізгі) – көмекші етістік (көмекші қызмет атқарады). Бұлар сыртқы тұрпаты жағынан да, ішкі мазмұны жағынан да құранды, мысалы: *дем ал, жақсы көр, үлгі қыл, ғалым бол*. Бұл келтірілген мысалдардың бірінші сыңары зат есім болса, екінші сыңары етістік түрінде құрастырылып, құрандалы тәсілмен жасалған.

2. Құрама күрделі етістік. Бұлардың әр компонентінің лексикалық мағынасы әртүрлі, бірақ әрбір тіркестің лексикалық мағынасы құрамындағы компоненттерінің мағынасынан құралады, бір күрделі мүше ретінде жұмсалатын болып қалыптасқан. Мысалы: *алып кел – ал да кел (ал мен келдің жеке тұрғандағы мағынасы бір-біріне ұқсамайды, әрбасқа); алып бар – ал да бар; алып кет – ал да кет*, т.б.

3. Сараламалы күрделі етістік. Бұл етістіктерді аналитикалық етістіктер деп атайды, етістік пен етістіктің тіркесінен тұрады. Алғашқы етістік негіз болады да, оған көсемше тұлғасы жалғанады, оған тіркесетін көмекші етістік негізгі етістіктің мағынасын саралайды, жіктейді, бірақ негізгі лексикалық мағынасын өзгертпейді. *Айта сал – не болса да айтып жібер* мағынасында, *айтып қал – қандай жағдай болмасын сөйлеп қал, айтып жібер – деп жібер, үндемей қалуға болмайды, айтып бер – сөйлеп бер, жауап бер* мағынасында, т.б. Яғни, әртүрлі мағыналарды саралауға көмекші етістік әсер етіп тұрғаны анық көрінеді.

4. Тұрақты күрделі етістік – идиомаланған тіркестер. Бұлар грамматикалық жағынан бөлшектенбейді, лексика-семантикалық жағынан бір ғана негізгі мағынаны білдіреді, бірбүтін тұлғаға айналған күрделі етістік түрі. Мысалы: *онық жеді – өкінді* мағынасында; *жан алқымға алды – қысты, көз салды – қарады, көңіл аударды*, т.б. [5, 239-240].

2002 жылы шыққан қазақ тілінің академиялық грамматикасында да А.Ысқақовтың классификациясы сипатталған [6]. Бұлайша топтаудың артықшылығы қазақ тілінде кездесетін күрделі етістіктердің барлығын қамтуға мүмкіндік береді. Бұл жерде айта кететін бір нәрсе, біз күрделі етістіктерді, яғни, күрделі сөздердің, олардың әр бөлек емес, бір мағынаны беретінін көрсетуді мақсат етіп отырғандықтан, жоғарыдағы төрт түрлі күрделі етістікке салт-дәстүр етістіктерін де қосуды жөн көрдік. Бұлар бір қарағанда құранды етістікке ұқсағанымен, мұның түзілуіндегі уәж басқа. *Салт-дәстүр етістіктері* аналитика-семантикалық тәсілмен жасалып, олардың негізгі мағынасы арқылы атауыштық жаңа мағына туындаған. Мысалы, төмендегі 1-кестеден салт-дәстүр етістігі мен тұрақты етістіктің мағынасы да, жасалу жолы да бір-бірінен бөлек екенін аңғаруға болады:

Салт-дәстүр етістігі мен тұрақты күрделі етістіктер

1-кесте

Салт-дәстүр етістігі	Тұрақты күрделі етістік
Қыз ұзату	Өңі кіру
Сырға салу	Мойнына алу
Келін түсіру	Ауыз жаласу
Сәлем салу	Ат құйрығын кесісу
Табак тарту	Төбе шашы тік тұру

Осындай ерекшеліктерін ескере келе, күрделі етістіктерді автоматты тануды жүзеге асыру үшін ең алдымен бес топқа жіктедік:

- *Құранды күрделі етістік;*
- *Тұрақты күрделі етістік;*
- *Салт-дәстүр етістігі;*
- *Құрама күрделі етістік;*
- *Аналитикалық формалы етістік.*

Негізгі мақсат бірнеше сөздің тіркесуі арқылы бір мағынаға ие болатын етістіктерді компьютерге бір сөз, бір токен ретінде таныту болғандықтан, осы топтар бойынша тілдік бірліктер жинақталып, олардың әрқайсына *Қазақ тілінің ұлттық корпусына* бейімделген аннотациялар берілді. Белгіленімдер 9 түрлі негізгі ақпаратты қамтыды:

- *лемма* – тіркескен күрделі етістіктің қосымшасыз формасы;
- *сөз табы* (етістік);
- *тұлғасы* (туынды);
- *құрамы* (күрделі);
- *сөзжасамдық тәсілі* (аналитикалық тәсіл);
- *күрделі етістік түрі* (құранды етістік, тұрақты етістік, салт-дәстүр етістігі, құрама етістік, аналитикалық формалы етістік);
- *моделі* (қай сөз таптарының қатысуы арқылы жасалғаны туралы ақпарат);
- *мағынасы* (күрделі етістіктің түсіндірмесі);
- *иллюстративті материал* (мысал).

ТАЛҚЫЛАУ

Сөз табы, тұлғасы, құрамы, сөзжасамдық тәсілі, күрделі етістік түрі сынды белгі-кодтар Қазақ тілінің ұлттық корпусында күрделі етістіктерге морфологиялық және сөзжасамдық белгіленімдерді қою үшін қажет. Бұл ҚТҰК-де күрделі етістіктерді танытудың алғашқы қадамы болмақ. ҚТҰК үшін дайындалып жатқан күрделі етістіктер тізімінде негізгі ақпараттың бір бөлігі ретінде лингвистикалық модельдер ұсынылады. Лингвистикалық модельдер күрделі етістіктердің қалай жасалғанын көрсете отырып, паттерндерді және сол паттерндерге тән мағыналық ерекшеліктерді айқындайды. Белгіленім-кодтың көрінісі келесідей (1-сурет):

Күрделі етістіктерді белгі-кодтау

1-сурет

	A	B	C	D	E	F	G	H	I	J
	Реестр	Лемма	Сөз табы	Тұлғасы	Сөз құрамы	Сөзжасамдық тәсіл	Етістік түрі	Мағынасы	Модель	Иллюстрация
										Мысал
1	Алып кел	алып кел	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Құрама етістік	бір затты немесе адамды бір орыннан екінші орынға әкелу	/етістік1+ып етістік2/	Бір соғым пұлсыз астан дейтіндер де Қасығын қолдарына а л ы п к е л д і (А.Байтұрсынұлы).
2	Шығып кет	Шығып кет	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Құрама етістік	жоғалтып алу	/етістік1+ып етістік2/	«Қолда барда алтынның қадірі жоқ, қолдан шығып кеткен соң ай өкіндім», – деп халық өленінде айтылатын ежелгі, дәстүрлі өлең табиғатына жанама нәр беріп, неге болса да тәубе етіп, неге болса да Құдай ісі деп, еш әрекетсіз қарап отыратын мінез-құлықтағы дертті аямайсыңа алады (Ө.Айтбаев, Ұлт мәдениетінің хантәңірі).
3	Тастап кет	Тастап кет	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Құрама етістік	артына қалдыру	/етістік1+ып етістік2/	Аханның алдында қазақ халқы үшін оның болашағы үшін өлшеусіз мол қазына т а с т а п к е т к е н үш алыптың – Абай, Ыбырай, Шоканның болғанын білеміз (Ө.Айтбаев, Ұлт мәдениетінің хантәңірі).
4	Қайтып кел	Қайтып кел	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Құрама етістік	Бір келген жеріне қайтадан келу	/етістік1+ып етістік2/	Ертегіне күн шыға күйеу жігіттер отырған бөлмеге бойдақ достарымен ақыры рет дузан құруға қ а й т ы п к е л д і (Ө.Кекілбаев, Мұнара). Осыдан кейін Евгений Цай журналистерге қ а й т ы п к е л м е д і (Г.Қапан, Ақиқатты айта бастап еді).
5	Кіріп шық	кіріп шық	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Құрама етістік	Бір жерге аз уақытқа кіріп қайтадан шығу	/етістік1+ып етістік2/	Үйіне көңіл айтуға к і р і п шы қ т ы (Ауызбекі).
6										
7										

Мұндағы күрделі етістіктердің *моделі* – күрделі етістерді танудың (суретте құрама етістіктің модельдері көрсетілген) әмбебап формуласы. Әр формуланың өзіне тән лексикалық мағынасы бар. Мәселен, *алып кел* құрама етістігінің *мағынасы* – бір затты немесе адамды бір орыннан екінші орынға әкелу. Ал оның *моделі* – **/етістік1+ып етістік2/**. Міне, мағына мен осы модель – бір токен боп саналады. Осы бір токен компьютерге үйретілетін бір материал болмақ.

Ал реестрде жинастырылған күрделі етістіктер тұтастай корпусқа дайын күйде беріледі де корпус оны бір тілдік бірліктер деп таниды. Бұлар – базаға кірген бірліктер саналады да *модель* өз кезегінде келешекте күрделі етістіктің осы типін (құранды етістікті) танудың, машинаны соған үйретудің бір тәсілі болмақ.

Күрделі етістіктердің ішінде тұрақты етістіктер мен салт-дәстүр етістіктеріне ғана мұндай алгоритмдер (модель) берілмейді. Себебі олардың мағынасы құрамындағы компоненттердің мағынасы және формасына тікелей тәуелді емес. Мысалы: «сәлем/салу» салт-дәстүр етістігі – біреу-ақ, оның ұқсас нұсқалары жоқ (2-сурет). Ендеше оның моделін түзіп қажет емес.

Салт-дәстүр етістігінің белгіленімі

2-сурет

1	Реестр	Лемма	Грамматика				Лексика	
			Сөз табы	Тұлғасы	Сөз құрамы	Сөзжасамдық тәсіл	Етістік түрі	Мағынасы
3	сәлем салу	сәлем сал	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	салт-дәстүр етістігі	инабаттылық пен әдепке қатысты қатынас нормасы.
4	ат қою	ат қой	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	салт-дәстүр етістігі	сәбиге есім беру, атау
5	айттыру	айт	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	салт-дәстүр етістігі	дәстүрлі ортада неке салтында үлкен мәнді құрып. Жігіт жағының құда түспек ниеттерін білдіріп, қыздың ата-анасына, жақын туыстарына сөз салуы.
6	ақ алып шығу	ақ алып шық	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	салт-дәстүр етістігі	ел бірлігі, ынтымақты мензейтін көшпелі ортада қалыптасқан моральдық этикалық ұғымды білдіретін құрып.

Құранды етістіктер, құрама етістіктер және етістіктің аналитикалық формаларын төмендегідей лингвистикалық модельдері әзірленді. Күрделі етістіктердің мағынасын көрсету олардың грамматикалық жақтан біртұтас тұлға ретінде танылуына ғана емес, семантикалық жақтан бір ұғымды білдіретінін де ажыратып береді.

Құранды күрделі етістік – есім және етістіктің тіркесуі арқылы жасалып бір ғана лексикалық мағынаны білдіретін сөздер тобы. Құранды күрделі етістіктердің құрамы да уақыт өте келе тұрақталып, біртұтастанған. Бұл жайында Ж.Шәкенов былай дейді: «Күрделі сөздердің өзге бір категорияға ауысуы немесе лексикалық тұлға болып қалыптасуы кездейсоқ болатын құбылыс емес, ауысу белгілі бір заңдылықтар негізінде жүзеге асады. Ондай заңдылықтар бірнешеу:

- лексикалану
- грамматикалану
- идиомалану

Мысалы, *той жасау, амал қылу, сый көрсету* сияқты басыңқы компоненті заттық ұғымды білдіріп, соған атау болатын сөздердің іс-әрекет, қимылды білдіретін сөздерге көшуі соңғы компонентте сөз тудыру қасиетінің болуынан» [7, 68]. Ғалым мұндай сөздердің лексикаланып тұрғанын мағына жағынан ұғымдас сөздердің тілімізде кездесуін дәлел ретінде ұсына отырып, былайша жіктеп көрсетеді:

Май басу – майлану
Шу көтеру – шулау
Жылт ету – жылтылдау
Сын тағу – сынау
Ілгері басу – ілгерілеу

Келтірілген мысалдарға қарағанда, сөздердің бір-бірінен айырмасы – жасалу тәсілінде екен. Бағананың бірінші қатарындағы сөздер аналитикалық тәсілмен жасалса, екінші қатардағы сөздер – заттық ұғымды жоғалтып, етістікке тән формалармен өзгеретін күйге жеткен сөздер, яғни бұл сөздер *грамматикаланған сөздер* болып табылады [8, 162]. А.Ысқақов осы тұлғалас күрделі етістіктер құрамындағы көмекші етістіктер беретін мағынасы жағынан (*шу көтеру – шу-лау*) жұрнаққа ұқсас болғанымен, формасы мен мағынасы жағынан ешуақытта онымен теңесе алмайтынын айтады.

М.Жолшаеваның пікірінше, құранды етістіктердің құрамындағы алдыңғы сыңарлары есім сөздер мен еліктеуіш сөздер оның (құранды етістіктің) лексикалық мағынасына негіз болады. Құранды етістіктердің құрамындағы көмекші етістіктердің сөзжасамдық қызметі туынды етістік жасайтын *-ла/-ле, -да/-де, -та/-те* жұрнағының қызметімен салыстыруға болады [9, 50].

Соған сай, кей құранды етістіктердің мағынасын оның бастапқы лексикалық мағынасымен (*жарқылдау, әңгімелесу, сөйлеу*) беруге тырыстық (3-кесте).

Құранды етістікке белгіленген мағыналар

3-кесте

<i>есіне түсу</i>	ұмытпай есіне сақтау	Қазіргі жолдасы Айдарбек болғандықтан, бұрын аралас-құраласы бар дос-жарандарының шеттеп кеткенін <i>есіне түсіреді</i> (Ғ.Мүсірепов, Белдерде).
<i>есіне алу</i>	ұмытпай есіне түсіру	Мәселе етіп қойды, көтерді, <i>есіне салды</i> . Америкада жүргенде Венгрия мәселесін, оқиғасын қайта-қайта атап, тың сұрау етіп алдарына көлденең тарта беріпті (М.Әуезов, Шығ.).
<i>жарқ ету</i>	жарқылдау	Пышағымды бауыздауға тақап тұрып, ұрып жіберейін дегенімде ажағай <i>жарқететүсті</i> (І.Жансүгіров, Шығ.).
<i>әңгіме қылу</i>	әңгімелеу	Әуезелеші айтатын сөзіне өзін қатыстырмай, өзінен тысқары ғаламда болған істі әңгіме қылады (А.Байтұрсынұлы, Шығ.).
<i>сөз сөйлеу</i>	сөйлеу	Келелі, назар аударарлықтай, көңіл бөлерліктей. – Іздеген қыздар табылды. Жоқты тапқан Күнтимес, Батыр Жолай алыптың Жалғыз қызы алымды, Ал, Күнтимес, <i>сөзсөйлеле</i> , Аты-жөнің танылды (К.Әзірбаев, Таңд. шығ.).

Құранды етістіктердің сыңарларын жеке-жеке бірлік ретінде қарастырсақ, мағыналары өзгеріп, қателікке ұрындыруы мүмкін. Алайда барлық құранды етістіктердің мағынасы біртұтастанып, күрделі ұғымға айналып үлгермегені белгілі. Қазақ мәтінін автоматты тану үшін күрделі етістіктің бұл түрін бір токенмен белгілеп, бір бірлік ретінде қарастыру мағынаны дұрыс анықтау үшін қажет.

Мысалы, құранды етістіктерді жасауда *қыл, бол, ет* етістіктері белсенді қолданылады (4-кесте):

Құранды етістіктердің лингвистикалық моделі (үлгі)

4-кесте

зат есім+қыл	<i>арман қылу, әдет қыл, ажуа қылу, әзіл қылу, дағды қылу, жақсылық қылу, жау қылу, жәрдем қылу, жыр қылу, келемеж қылу, қанағат қылу, қонақ қылу, қызмет қылу, мазақ қылу, риза қылу, ойран қылу, сәлем қылу, сабыр қылу, сылтау қылу, өсек қылу, талқан қылу, үміт қылу</i> т.б.
сын есім+қыл	<i>алаң қылу, әбігер қылу, әуре қылу, қор қылу, құмар қыл, құштар қылу</i> т.б.

Ізденісіміздің барысында 200-ге жуық құранды етістік жиналып, мағынасы ашылып, модельдері қойылды.

Құрама етістіктер – күрделі етістіктердің басқа түрлеріне қарағанда мағынасы да, құрамы да тұрақтанбаған сөздер. Қазақ тіліндегі кез келген мағынасы жағынан үйлесімді екі

әрекет көсемше жұрнағы (-ын/in/n, -а/е/й) арқылы бір-бірімен байланысып құрама етістіктер қалыптасады. Сол себепті олардың нақты санын анықтау, толық тізімдеп шығу тіптен мүмкін емес. Сондай-ақ олар сыртқы тұрпаты жағынан аналитикалық формалы етістіктерге, еркін сөз тіркестеріне ұқсас келеді (5-кесте).

Құрама етістіктер мен сөз тіркестері

5-кесте

Етістік	Түр (вид)	Моделі
Ұшып тұр	Құрама етістік	етістік₁+ып-етістік₂
Ұшып өт	Еркін сөз тіркесі	-
Ұшып кір	Еркін сөз тіркесі	-
Ұшып кет	Еркін сөз тіркесі	-
Ұшып бар	Еркін сөз тіркесі	-

Құранды етістіктердің басты белгісі құрамындағы етістіктердің мағынасы сақтала отырып, жаңа бір күрделі әрекетті білдіретін мағына туындайды. Бұл жайлы Ы.Маманов былай дейді: «...мұнда (күрделі етістік – ред.) негізгі етістіктер де, көмекші етістіктер де өз мағынасында тұрып, бір тұтас ойды білдіретін сөз тіркесін жасайды. Бұлардың кейбірі ықшамдалып, өзара бірігіп, бір сөзге айналып кетеді: алып бар – апар, алып кел – әкел, алып кет – әкет» [10, 61].

Салт-дәстүр етістіктері және тұрақты күрделі етістіктер. Корпус базасында танылмайтын тұрақты етістіктер мен салт-дәстүр етістіктерінің тізімі жасалып, 400 қолданыс жиналды. Автоматты тану процесінде ауытқулар болмас үшін, етістіктердің грамматикалық формаларының өзгерістері, яғни варианттары да осы тізімге қосылды. Атап айтқанда, төмендегідей үлгімен фразеологиялық варианттар қамтылды:

- *Басым қатты* – *басы қатты* – *басың қатты*;
- *Қолыма түсті* – *қолына түсті* – *қолға түсті*;
- *Есесі кетті* – *есем кетті* – *есеміз кетті*

Грамматикалық тұлғалар, яғни жалғаулар бірінші сөзден кейін жалғанатындықтан, «бас қат», «қолға түс», «есе кет» сынды негізгі тұлғаларын ғана тізімге енгізсек, жүйе *басы қатты*, *қолына түсті*, *есеміз кетті* сынды қолданыстарды танымауы мүмкін. Осы мәселе тізімдеу барысында ескерілді.

Этнографиялық ақпараты мол салт-дәстүр етістіктері көркем әдебиет стилінде, публицистикалық стильде көп кездеседі. Ал аталған стильдерде жазылған мәтіндер корпус базасының негізін құрайды. Сол себепті салт-дәстүр етістіктерін бір ұғым ретінде белгілеу қолданушының аялық білімін кеңейтуге ықпал етеді. Жиналған тізімдегі салт-дәстүр етістіктерінің мағынасы қысқаша ашылған (6-кесте).

Салт-дәстүр етістіктеріне белгіленген мағыналар

6-кесте

Ашамайға мінгізу	Ер бала 4-5 жасқа келгенде бәсіре тайына ашамай ерттеп, оны алғаш рет атқа мінгізу рәсімі.
Ауыз бекіту	Ораза тұтушының таң намазы уақытынан бұрын тамақтанып, А. б.-ге ниет еткен діни сүннеті.
Ат тұлдау	Дүниеден озған ер адамның мінген атының жал-құйрығын шорт кесіп күзеп тастау рәсімі.
Бақан салу	Киіз басқанда тізгін жіптің орнына жүнге тигізбей шидің басына екі басынан шығып тұратын бақанды салып, орап шиыршықтау.
Құда түсу	Біреудің қызын баласына айттыра бару (келу).

Қамшы қалдыру	Құда түсу салтында қыз көруге келген адамдардың құда болуға ниеті бар екенін білдірудің тілдік емес (бейвербалды) белгісі.
Бие байлау	Қымыз ашыту үшін желіге құлынды байлап, бие сауу.
Аяқ жүгірту	Құда түсу салтында қыз көруге келген адамдардың құда болуға ниеті бар екенін білдірудің тілдік емес (бейвербалды) белгісі.

Салт-дәстүр етістіктердің атаулық мағынасы басым болғанымен, мәтіндегі қолданысқа келгенде етістік позициясына көшетіні байқалады: *ашамайға мінгізді, ауыз бекітті (ауызын бекітті), ат тұлдады, бақын салды, құда түсті*, т.б. Сондай-ақ, кейде мұндай етістіктердің арасында қосымшалар түсіп кететінін байқауға болады: *қыз ұзатты – қызын ұзатты – қыздарын ұзатты, қызымды ұзаттым*, т.б. Осындай грамматикалық ерекшеліктеріне қарап мұндай етістіктерді де күрделі етістіктердің құрамында бөлек топ ретінде қарауға болады деген ойдамыз.

Аналитикалық формалы етістіктер. Қазақ тіліндегі автоматтандыруға, машинаны үйрету процестеріне ең ыңғайлы тілдік бірліктердің бірі – етістіктің аналитикалық форманттары. Аналитикалық формант (*ары қарай – АФ*) дегеніміз – қосымша мен көмекші етістіктің бірлігінен тұрып, грамматикалық мағына беретін морфологиялық көрсеткіш. Мәселен, *айтып бер* деген күрделі етістікті алып көрейік. Мұнда *бер* етістігі бір нәрсені беру, яғни, ұстату, қол жалғау секілді лексикалық мағыналарды беріп тұрған жоқ. Яғни, әу бастағы лексикалық мағынасын жоғалтып, бағыныңқы сыңарға айналған. Ал басыңқы бірлік – *айтып* етістігі өзінің негізгі мағынасында тұр. Ал екі тілдік бірлік *-ып* көсемшесі арқылы дәнекерленіп тұр. Осы *-ып бер* дәнекерін – аналитикалық формант дейміз. Ал оның үстеп тұрған мағынасы – қимылды толық аяқтауды білдіру. Бұл етістіктерді бір тұлға етіп көрсетіп тұрған көсемше, есімше, рай жұрнақтары мен көмекші сөздер. Схемасы мынадай болады:

Аналитикалық етістік = негізгі етістік + дәнекер форма + көмекші етістік айт (негізгі етістік) + **ып** (дәнекер форма немесе формант) + **бердім** (көмекші етістік).

Ағылшын тіліндегі фразалық етістіктерге (*get up, come up, go up, sit up, talk up*) ұқсас мұндай аналитикалық форманттар да жинақталып, оларды машинаны үйретуге бір токен ретінде беру үшін мағыналарын ашып, жүйелеуіміз керек. Қазақ тіліндегі етістіктердің аналитикалық форманттары генеративті ЖИ-ді үйретуде маңызды рөл атқарады. Өйткені олар етістіктің шақ, рай, модальділік және аспектуалды сипаттарын анықтайды. Қазақ тілінің аналитикалық форманттарын дұрыс түсіну модельдің сөйлемнің грамматикалық құрылымын және мағыналық реңктерін дәл анықтауына көмектеседі.

Қазақ тіл білімінде аналитикалық формалар туралы алғаш сөз қозғаған – А.Ысқақов. Оның «Қазіргі қазақ тілі. Морфология» кітабының алғашқы 1963 жылғы басылымында аналитикалық форма туралы сөз болмағанмен, кейінгі 1974 жылғы екінші басылымында «Көмекші сөз бен аналитикалық форма», «Аналитикалық форманттар және олардың мағыналары» деген бөлімдерін қосқан [11].

Одан кейінгі 1975 жылы жарық көрген Н.Оралбаеваның «Қазіргі қазақ тіліндегі етістіктің аналитикалық форманттары» деген еңбегі осы мәселеге толықтай арналған алғашқы еңбек саналады. Ғалым АФ-лар толық мағыналы етістікке тіркеседі әрі етістіктің белгілі бір категорияларының қосымша мағыналарын беруші көрсеткіштер деп көрсеткен [12, 3]. Ал өз кезегінде С.Исаев аналитикалық форманттарды күрделі етістіктің құрамында қарауды жөн деп санайды да модаль мәнді сөздермен тіркесіп келгендерді де осы аналитикалық формант құрамында қарастыру қажетін айтады [13, 251].

АФ-лардың морфемдік құрамы тұрақты, құрылымы қалыптасқан, арасына басқа бір тілдік бірліктер түсіп кетпейді. Мәселен, *-а бер, -е бер, -й бер, -и бер, -ып таста, -іп кеп бар*, т.б. Әр аналитикалық форманттың өзіне тән грамматикалық мағынасы бар, етістікке тіркесу арқылы толық мағыналы етістікке қосымша грамматикалық мағыналар

үстейді. Мұндай құрылымы анық, мызғымас, тұрақты тілдік бірліктерді кодтау – тілді автоматтандырудың бірден-бір ұтымды жолдарының бірі. Оның үстіне, қазақ тілі үшін күрделі етістіктердің маңызы зор. Күрделі етістіктер – қазақ тілінің грамматикасына тән, көпфункционалды қызмет атқаратын формалар. Ал күрделі етістерді жасаушы формалардың бірі – аналитикалық форманттар.

Жоғарыда айтылғандай, әр аналитикалық форманттың өзіне тән грамматикалық мағынасы болады. Мәселен, *-а жазда* аналитикалық форманты арқылы жүзеге аспай қалған қимылды, *-а бер* – қимылдың қайталануын, *-а қал* – қимылдың аяқ астынан, жеңіл жасалуын, *-п таста* – қысқа, батыл жасалған қимылды білдіреді.

Н.Оралбаева қазақ тіліндегі АФ-лардың грамматикалық мағынаны білдіруіне қарай екі үлкен топқа бөледі:

1. Қимылдың жасалу тәсілдерін білдіретін АФ-лар;
2. Қимылдың жасалу кезеңін (басталу, жасалу, аяқталу) білдіретін АФ-лар.

Сөйтіп, ғалым 60 аналитикалық форманттың 200-ге жуық нұсқасының (инвариант) 40 түрлі формасын көрсеткен. Біздің ізденісімізде 328 аналитикалық форманттың (варианттары мен инварианттары) 238 моделі тұрғызылды, яғни, типтік мағынасы ашылды (3-сурет).

АФ-ларға берілген белгіленімдердің үлгісі

3-сурет

1	Реестр	Форма	Сөз табы		Грамматика		Моделі	Лексика	Иллюстрация
			Сөз табы	Сөз құрамы	Сөзжасамдық тәсіл	Етістік түрі		Мағынасы	Мысал
2	ішіп салды	-ып/ін/п сал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік + /н/ сал	қимылдың өту сипатын, оның ішінде жеделдігін білдіреді	Жеке тақ жолдасының тілін алды. Шапанын шешті дағы, төсе п салды (А.Байтұрсынұлы). Екеуі жымды судан ішіп п салды. Құтырып көп алдына жетіп барды (Абай).
3	айтқан болды	-қан бол	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+қан бол	қимылдың жасалуындағы жасандылықты білдіреді	Біраз адам ол аялдаманың бар екенін, өзге маршруттар тоқтап жүргенін, арасында осы 73-ге тоқтайтынның айтқан болды (М.Шамұратов)
4	жасырған болды	-ған бол	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+ған бол	қимылдың жасалуындағы жасандылықты білдіреді	Ол қатұланып отырса да, қызметшінің бұл ағаттығын көшірген болды, лықсып келе қалған ашуын жасырған болды (Ө.Кекілбаев)
5	кеткен болды	-кен бол	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+кен бол	қимылдың жасалуындағы жасандылықты білдіреді	Ауыз үйге кеткен болды (Ауызекі тіл).
6	кешірген болды	-ген бол	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+ген бол	қимылдың жасалуындағы жасандылықты білдіреді	Ол қатұланып отырса да, қызметшінің бұл ағаттығын кешірген болды, лықсып келе қалған ашуын жасырған болды (Ө.Кекілбаев)
7	аудары салды	-а сал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+а сал	қимылдың немқұрайлы жасалуын білдіреді	Жұмыртқа келіншек ерінін сылып еткізіп қалта тапқан қаһармандай дорбаны столға аудары салды (Қ.Өзденбаева). Адал жолға салды.
8	бере салды	-е сал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+е сал	қимылдың немқұрайлы жасалуын білдіреді	Қыз бала салды қырықты (Е.Дүйсенбайұлы). Өзім төлеймін, деп қалтасынан бұрқыраған бір бұма мыңдықтардың бірін алып, қызға берсе салды (Қ.Өзденбаева)

Мәселен, 3-суретте көрсетілгендей, *-қан бол* аналитикалық формантының үш фонетикалық варианты бар: *-ған бол*, *-ген бол*, *-кен бол*. Бұлардың білдіретін мағынасы – қимылдың жасалуындағы жасандылық. Бұл – қимылдың жасалу тәсілі. Бұл бір модель, яғни бір токен боп танылады. Яғни, *-ған бол* аналитикалық форманттары арқылы жасалған күрделі етістердің барлығы қимылдың жасалуындағы жасандылықты білдіреді деген мағына үйретіледі компьютерге.

Сондай-ақ, бір мағыналы аналитикалық форманттармен қатар көпмағыналы АФ-лар да бар. Модельде бұлардың екі мағынасы да көрсетілді (4-сурет).

Көпмағыналы АФ-лар

4-сурет

үстай алды жаза ала ма? айта ала ма?	-а ал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+а алмай қал	1. қимылдың бірден, тез жасалуын көрсетеді; 2. қимылды жасауға объектінің мүмкіндігі бар-жоғын білдіру	1. Абай терезе алдынан атып тұрып, Оспанның анаған бір зорлық қылып келе жатқанын сезіп, алдынан шығып ұстағайлады (М.Әуезов). 2. Бұрын фактологиялық білімді, грамматиканы тексерген болсақ, енді бала оқи ала ма, жаз ала ма, айта ала ма, түсіндіре ала ма? Соны тексеретін болдық (Р.Рақымықызы).
оқыта беріңіздер тіке тарта беріңдер	-а бер	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+а бер	1. қимылдың қайталануы мағынасын беруде қолданылады 2. қимылдың басталғанын білдіреді	1. Балаларды оқыта беріңіздер (М.Ақмырзаұлы). 2. Мені күтпей, аэродромға тіке тарта беріңдер (Ө.Таразы).

Көбіне аналитикалық форманттар екі компонентті болып келетіні белгілі. Дәнекер форма көсемше не есімше тұлғасында тұрады да оған келіп көмекші етістік тіркеседі. Бірақ қазақ тілінде үш, төрт компонентті аналитикалық форманттар да кездеседі.

Мәселен, *-п қоя бер, -п сала бер, -п келе жат, -п бара жат, -п кеп жібер, -п жүре бер, -п ала қой, -п шыға кел, -п кеп жібере жазда* – күрделенген аналитикалық форманттар (5-сурет).

Көпкомпонентті АФ-лар

5-сурет

бөлейін деп отыр	-йін деп отыр	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+йін деп отыр	қимылдың жасалуға бейім тұру мағынасы	«Ақмоладосервис и К» ЖШС бұған қосымша техниканы бөлейін деп отыр (А.Қалжанов).
айтайын деп отырмын	-йын деп отыр	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+йын деп отыр	қимылдың жасалуға бейім тұру мағынасы	Басымнан өткен оқиғамды айтайын деп отырмын (Ауызекі тіл).
жазып кеп қалды	-ып/-іп/-п кеп қал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+п кеп қал	ерекше қарқынмен жасалған қимыл мағынасын береді	Жазып кеп қалды жиырылған қасын Күлімдеп қария қадала берді Ұнатып қара көзімен шашын (Ө.Тәжібаев).
Айта алмай қалды	-а алмай қал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+а алмай қал	істің орындалмай қалғанын білдіреді	Жетті де байғұз дегенге, Айта алмай қалды ар жағын Килігіп кенет бегеуге, Батырды күткен байғұз құс Білгенін айтып болыпты (Ө.Тәжібаев).
сайрап кеп берді	-ып/-іп/-п кеп бер	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+п кеп бер	ерекше қарқынмен жасалған қимыл мағынасын береді	Күлжәмиле тағы сайрап кеп берді (К.Найманбаев).

Сондай-ақ, жоғарыдағы мысалда көрсетілген *-а алмай қал* моделі – *-а ал қал* моделінің бір варианты болғанымен, біздің ойымызша, компьютер тілдегі барлық АФ-ларды тану үшін олардың болымсыз формаларын да қатар беру қажет. Сондықтан ондай формаларды жинастыру қажет деп білеміз. Жоғарыда соның үлгісі көрсетілген.

Н.Оралбаева аналитикалық форманттарының арасында кейде демеулік шылаулардың түсіп кететінін айтқан [12, 44]. Мәселен, *Кейде жарасып та кетеді* (М.Елбаев) деген сөйлемде *-п та кет* демеулік шылаулы АФ тұр. Автоматтандыруды белгілеуде мұны да қамтуға тырыстық (6-сурет).

Демеулік шылауы бар АФ-лар

6-сурет

жарасып та кетеді	-ып/-іп/-п та кет	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+ /п/ та кет	қимылдың бірден, тез жасалуын көрсетеді	Кейде ол жарасып та кетеді (М.Елбаев).
өліп те кетеді	-ып/-іп/-п те кет	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+ /п/ те кет	қимылдың бірден, тез жасалуын көрсетеді	Кейде өліп те кетеді (Г.Жұмабайқызы).
оқи да біледі	-и да біл	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+и да біл	қимылды жасауға объектінің мүмкіндігі бар-жоғын білдіру	Өзіндей балалар оқу программасын біраз еңсеріп тастады, алды аздап жаз да, оқи да біледі (Д.Өшмісханұлы).
жаза да біледі	-а да біл	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+а да біл	қимылды жасауға объектінің мүмкіндігі бар-жоғын білдіру	Ағылшынша оқи да, жаз да біледі.

Белгіленім жұмысында етістіктің тұрақты аналитикалық форманттардан бөлек құрамы мен құрылысы аналитикалық форманттарға ұқсайтын, бірақ мағынасында категориялық грамматикалық мағыналар үстелетін АФ-ларды да қамтуға көңіл бөлінді (7-сурет). Бұларды кейде модальдық мағынаны беретін аналитикалық формалар деп те атайды.

Модальді мағынаны білдіретін АФ-лар

7-сурет

болғысы келеді	-ғысы кел	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+ғысы кел	қалау мәнді білдіреді	... "Нұр Отан" елдің саяси аренасында жалғыз болғысы келеді (ЖА).
бергісі келмей қалды	-гісі келмей қал	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+гісі келмей қал	қалау мәнді білдіреді	«Император Путин» билікті бергісі келмей қалды (Abai.kz).
қайтқалы жүр	-қалы жүр	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	басталғалы отырған іс-әрекетті білдіреді	Құжаттарын жинастырып, қайтқалы жүр дегенді есітіп: – Ау-у, мұныыз не? – дедік танырап. (Ш.Дәрмағамбетұлы).
баратын шығар	-атын шығар	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+атын шығар	шарттың келер шақта орындалуын білдіреді	Биыл бір жақта шығатын болсақ, не Бурабайға, не Алакөлге баратын шығармыз (Азаттық Радиосы).
үзілетін шығар	-етін шығар	Етістік	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік+етін шығар	шарттың келер шақта орындалуын білдіреді	Емі тезірек табылмаса, кешікпей үзілетін шығар (Ө.Өбішев).

Бұлар қазақ тілінің грамматикасында қалыптасып қалған аналитикалық форманттарға жатпаса да, сыртқы формасы мен ішкі мазмұны АФ-ға ұқсас болғандықтан, біз де мұны АФ-лар деп танып, белгіленім жұмыстарын жүргіздік.

Аналитикалық форманттардың сыртқы формасы күрделі етістіктерге, әсіресе, құрама етістіктерге өте ұқсас. Сондықтан белгіленімде оның аражігін ажырату аса маңызды. Мәселен, 8-суреттегі мысалда «төсеп салды» мен «ішіп салды» күрделі етістіктерінің сыртқы формасы бір, яғни, *-п сал* аналитикалық формантына ұқсас.

АФ мен құрама етістік

8-сурет

1	A	B	C	D	E	F	G	H	I	J
2	Реестр	Форма	Сөз табы	Тұлғасы	Сөз құрамы	Сөзжасамдық тәсіл	Етістік түрі	Модель	Лексика	Иллюстрация
3	ішіп салды	-ып/ін/п сал	Етістік	Туынды	Күрделі	Аналитикалық тәсіл	Аналитикалық етістік	етістік + /п/ сал	Мағынасы	Мысал
									қимылдың өту сипатын, оның ішінде жеделдігін білдіреді	Жеке тақ жолдасының тілін алды, Шапаның шешті дағы, <u>төсеп салды</u> (Абайтұрысқұлы). Екеуі жаныз суан <u>ішіп салды</u> . Құтырып көп алдына жетіп барды (Абай).

Бірақ бұл жерде «*төсеп сал*» күрделі етістігінің екі лексемасы да лексикалық мағынаға ие, ал «*ішіп сал*» күрделі етістігінде *сал* көмекші етістігінің мағынасы солғын, яғни, грамматикалық мағына үстеп тұрған бірлік екенін байқауға болады. Аналитикалық форманттың құрамындағы көмекші етістік толық лексикалық мағынаға ие болмайды, тек грамматикалық мағына үстеуші тілдік бірлік саналады. Қазақ тілінің ұлттық корпусындағы мұндай ұқсас тілдік бірліктерді ажырату үшін – оларды әрқайсын тауып, АФ ретінде бөлек, құранды етістік ретінде бөлек белгіленім жүргізу қажет. АФ-мен байланысты алдағы жұмыстар осы бағыттарда жүргізілмек.

ҚОРЫТЫНДЫ

Күрделі етістіктер – грамматикалық, модальдік және семантикалық, аспектуалдық, басқа да жұмсалымды мағыналарды білдіретін қазақ тіліндегі бай, күрделі бірліктер. Оны машинаға үйретудің дұрыс бағытын таңдап, мультимодальды корпус жүйесінде морфологиялық, синтаксистік және семантикалық деңгейлерде аннотациялау жұмыстарын жүргізу өте маңызды. Бұл жұмыстар үлкен тіл модельдерінің қазақ тіліндегі дәлдігін арттырып, қазақ тілінің цифрлық дамуына үлкен үлес қосады. Сол арқылы үлкен тіл модельдері қазақ тілін тереңірек түсініп, адам дүниетанымындағы басты логикалық категориялардың бірі – іс-қимылды ажыратудың амалдарын меңгеріп, сауатты қазақтілді жасанды интеллект жасауға жол ашпақ.

Күрделі етістіктердің ішінде аналитикалық форманттар арқылы жасалатын күрделі етістіктер модельдеу мен тұрақты күрделі етістіктерді тегтеу біршама оңайырақ екенін, ал құрама етістіктерді тегтеу модельдеуге келгенімен, көп уақыт алатыны, ал құранды етістіктерді тегтеу де біршама уақыт пен зерттеу жұмыстарын талап ететіні белгілі болды. Дегенмен күрделі етістіктерді модельдеудің біршама тәжірибесі қалыптасты деуге келеді, кейбір тереңірек жүргізілетін машинаны үйрету процестеріне әлі де грамматикалық зерттеу жұмыстары қажет екенін анықталды. Алдағы корпусстық қолданбалы жұмыстар осы бағыттарда дамуы керек деген сенімдеміз.

ҚАРЖЫЛАНДЫРУ ТУРАЛЫ АҚПАРАТ

Мақала «Қазақ тілі ұлттық корпусын Smart-мәтіндер мегажобасы және қазақтілді жасанды интеллект негізі ретінде жетілдіру, ішкорпустарын әзірлеу» (BR24993244) бағдарламалық-нысаналы қаржыландыру жобасы аясында жазылды.

ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Toporkov O., Agerri R. On the Role of Morphological Information for Contextual Lemmatization // Computational Linguistics. – 2024. – №50 (1): 157–191 pp. https://doi.org/10.1162/coli_a_00497
- 2 Oxford Phrasal Verbs Dictionary for learners of English. Oxford University Press, 2006. – 436 p.
- 3 Slyambekov Q.B., Sadyk A.M. The National corpus of the Kazakh language: the semantic markup of verbs // Tiltanym, 2024. – №1. – 189-196 pp. <https://doi.org/10.55491/2411-6076-2024-1-189-196>

- 4 Toluspayeva, G. The System of Grammatical Categories of the Verb in Kazakh, Russian, and English // World Journal of English Language, 2023. – №13(2). – 120-126 pp. <https://doi.org/10.5430/wjel.v13n2p120>
- 5 Ысқақов А. Қазіргі қазақ тілі. Морфология. – Алматы: Ана тілі, 1991. – 381 б.
- 6 Қазақ грамматикасы. Фонетика, сөзжасам, морфология, синтаксис. – Астана, 2002. – 784 б.
- 7 Шәкенов Ж. Қазақ тіліндегі күрделі сөздер мен күрделі тұлғалар. – Алматы: Ана тілі, 1991. – 120 б.
- 8 Момынова Б. Қазақ тілінің морфологиясы: оқу құралы. – Алматы: Қазақ ун-ті, 2014. – 278 б.
- 9 Жолшаева М. Қазақ тіліндегі құранды етістіктердің лексика-грамматикалық ерекшеліктері// Tiltanym. 2022. – №1 (185). – 44-54 pp. <https://doi.org/10.55491/2411-6076-2022-1-43-52>
- 10 Маманов Ы. Қазіргі қазақ тілі: Етістік. – Алматы: Мектеп, 1966. – 156 б.
- 11 Ысқақов А. Қазіргі қазақ тілі. Морфология. – Алматы: Мектеп, 1974. – 407 б.
- 12 Оралбаева Н. Қазіргі қазақ тіліндегі етістіктердің аналитикалық форманттары. – Алматы: Мектеп, 1975. – 136 б.
- 13 Исаев С. Қазіргі қазақ тілі. Морфология. – Алматы: Өнер, 2007. – 336 б.

Материал 03.01.2025 баспаға түсті

Опыт обучения систем искусственного интеллекта сложным глагольным конструкциям казахского языка: подходы к тегированию и морфологическому моделированию

К.Б. Слямбеков¹, А.А. Солтанбекова¹, Б.С. Шалабай², Т. Калибекулы³

¹Институт языкознания им. А. Байтурсынова, Алматы, 050040, Казахстан

²Кокшетауский университет имени Ш. Уалиханова, Кокшетау, 020000, Казахстан

³Казахский университет международных отношений и мировых языков имени Абылай хана, 050022, Алматы, Казахстан

В статье изложен опыт реализации этапов аннотирования и тегирования для построения генеративных моделей ИИ посредством автоматического распознавания сложных глаголов в казахском языке и проведения работ по разметке корпуса. Рассматриваются трудности и научные проблемы при тегировании сложнокомпонентных глаголов, составных глаголов и аналитических глагольных форм, пути их решения, а также особенности разметочных процедур. Несомненно, в основе архитектуры больших языковых моделей лежит грамматика соответствующего языка, и его грамматические модели. Направление развития современной лингвистики также связано с разработкой и совершенствованием таких моделей. Поэтому моделирование одного из сложных грамматических компонентов казахского языка — сложно-составных глагольных конструкций — и предложение их морфологической схемы разметки является одной из наиболее актуальных задач автоматизации казахского языка. Настоящая статья написана с целью поделиться опытом, накопленным в этом направлении.

Ключевые слова: глагол, сложные глаголы, аналитический формант, составные глаголы, правильные сложные глаголы, искусственный интеллект, автоматизация, разметка

СПИСОК ЛИТЕРАТУРЫ

- 1 Toporkov O., Agerri R. On the Role of Morphological Information for Contextual Lemmatization // Computational Linguistics. – 2024. – №50 (1): 157–191 pp. https://doi.org/10.1162/coli_a_00497
- 2 Oxford Phrasal Verbs Dictionary for learners of English. Oxford University Press, 2006. – 436 p.
- 3 Slyambekov Q.B., Sadyk A.M. The National corpus of the Kazakh language: the semantic markup of verbs // Tiltanym, 2024. – №1. – 189-196 pp. <https://doi.org/10.55491/2411-6076-2024-1-189-196>

- 4 Toluspayeva, G. The System of Grammatical Categories of the Verb in Kazakh, Russian, and English // World Journal of English Language, 2023. – №13(2). – 120-126 pp. <https://doi.org/10.5430/wjel.v13n2p120>
- 5 Ысқақов А. Современный казахский язык. Морфология. – Алматы, 1991. – 381 с.
- 6 Казахская грамматика. Фонетика, словообразование, морфология, синтаксис. – Астана, 2002. – 784 с.
- 7 Шакенов Ж. Сложные слова и сложные лица в казахском языке. – Алматы, 1991. – 120 с.
- 8 Момынова Б. Морфология казахского языка: учебное пособие. – Алматы: Казахский ун-ті, 2014. – 278 с.
- 9 Жолшаева М. Лексико-грамматические особенности составных глаголов в казахском языке // Tiltanym. 2022. – №1 (185). – 44-54 pp. <https://doi.org/10.55491/2411-6076-2022-1-43-52>
- 10 Маманов Ы. Современный казахский язык: Глагол. – Алматы: Мектеп, 1966. – 156 с.
- 11 Ысқақов А. Современный казахский язык. Морфология. – Алматы: Мектеп, 1974. – 407 с.
- 12 Оралбаева Н. Аналитические форманты глаголов в современном казахском языке. – Алматы: Мектеп, 1975. – 136 с.
- 13 Исаев С. Современный казахский язык. Морфология. – Алматы: Онер, 2007. – 336 с.

Материал поступил в редакцию журнала 03.01.2025

Experience in Training AI Systems on Complex Verb Constructions in Kazakh: Approaches to Tagging and Morphological Modeling

K. Slyambekov¹, A. Soltanbekova², B. Shalabai³, T. Kalibekuly⁴

¹The Institute of Linguistics named of after A. Baitursynov, Almaty, 050040, Kazakhstan

²Sh. Ualikhanov Kokshetau University, Kokshetau, 020000, Kazakhstan

³Kazakh Ablai Khan University of International Relations and World Languages, Almaty, 050022, Kazakhstan

The article presents practical experience in implementing the stages of annotation and tagging for building generative AI models through the automatic recognition of complex verbs in Kazakh and corpus annotation work. It discusses the difficulties and scholarly issues involved in tagging multi-component verbs, compound verbs, and analytic verbal forms, the ways to address them, and the specific features of the annotation procedures. It is beyond doubt that the grammar of the language underlies the architecture of large language models, and that their transformations are grounded in the language's grammatical models. The trajectory of modern linguistics is likewise tied to the development and refinement of such models. Therefore, modeling one of the complex grammatical components of Kazakh – complex verb constructions – and proposing their morphological annotation scheme is among the most pressing tasks in the automation of the Kazakh language. This article has been written to share the experience accumulated in this area.

Key words: verb, compound verbs, analytical formant, compound verbs, regular compound verbs, artificial intelligence, automation, markup

REFERENCES

- 1 Toporkov O., Agerri R. (2024), "On the Role of Morphological Information for Contextual Lemmatization", *Computational Linguistics*, 50 (1), pp. 157-191.
- 2 Oxford Phrasal Verbs Dictionary for learners of English. Oxford University Press, Oxford, UK.
- 3 Slyambekov Q.B., Sadyk A.M. (2024), "The National corpus of the Kazakh language: the semantic markup of verbs", *Tiltanym*, 1 (93), pp. 189-196.
- 4 Toluspayeva, G. (2023) "The System of Grammatical Categories of the Verb in Kazakh, Russian, and English", *World Journal of English Language*, 13 (2). – 120-126 pp. <https://doi.org/10.5430/wjel.v13n2p120>

- 5 Ysraqov A. (1991) *Qazirgi qazaq tili. Morfologija* [Contemporary Kazakh Language. Morphology], Almaty. (in Kazakh)
- 6 Qazaq gramatikasy. *Fonetika, sözjasam, morfologia, sintaksis* [Kazakh grammar. Phonetics, word formation, morphology, and syntax], Astana. (in Kazakh)
- 7 Shakenov Zh. *Qazaq tilindegi kurdeli sozder men kurdeli tulgalar* [Compound Words and Complex Forms in the Kazakh Language], Almaty. (in Kazakh)
- 8 Momynova B. *Qazaq tiliniñ morfologiasy: oqu qūraly* [Morphology of the Kazakh language: manual], Almaty. (in Kazakh)
- 9 Zholshaeva M. (2022), *Qazaq tilindegi qurandy etistikterding leksika-grammatikalyq erekshelikteri* [Lexico-grammatical features of compound verbs in Kazakh] // Tiltany. – 2022. – Vol.1 (85). – pp. 44-54.
- 10 Mamanov Y. *Qazirgi qazaq tili. Etistik* [Morphology of the Kazakh language. Verb], Almaty. (in Kazakh)
- 11 Ysraqov A. (1974) *Qazirgi qazaq tili. Morfologija* [Contemporary Kazakh Language. Morphology], Almaty. (in Kazakh)
- 12 Oralbaeva N. (1975) *Qazirgi qazaq tilindegi etistikterding analitikalyq formantтары* [Analytical formatives in Contemporary Kazakh], Almaty. (in Kazakh)
- 13 Isaev S. *Qazirgi qazaq tili. Morfologia* [Modern Kazakh language. Morphology], Almaty. (in Kazakh).

Received: 03.01.2025

MRNTI 16.21.49

DOI: [10.59102/kufil/2025/iss3pp108-122](https://doi.org/10.59102/kufil/2025/iss3pp108-122)

S. Tagirova¹, Z. Khabibullina¹, A. Kurmanova², N. Smagulova²

¹Bashkir State Pedagogical University named after M. Akmulla, Ufa, 450000, Russia

²Kokshetau University named after Sh. Ualikhanov, Kokshetau, 020000, Republic of Kazakhstan

COMPARATIVE ANALYSIS OF PHRASEOLOGISMS IN THE BASHKIR AND KAZAKH LANGUAGES

The article is devoted to the phraseology of the Bashkir and Kazakh languages as a reflection of the national mentality and linguistic worldview. The relevance of the study is due to the need for a deeper understanding of the general and specific features of phraseological units of Turkic languages in the context of intercultural interaction. The aim of the study is to conduct a comparative analysis of phraseological units of the Bashkir and Kazakh languages, to identify their structural and semantic features, and to determine their role in the transmission of cultural and historical values of these peoples. The work is based on the methods of comparative analysis, component and semantic analysis, as well as a linguacultural approach. The use of these methods made it possible to identify both general patterns and differences in the functioning of phraseological units of the two languages, which are explained by their common Turkic heritage and historical interaction. At the same time, differences related to the cultural traditions and lifestyle of each ethnic group are identified. Expressions denoting hard work are represented in the Bashkir language by means of phraseological unit "altyn qullı" (with golden hands), and in the Kazakh language by "barmağynan bal tamğan" (honey drips from the finger), which reflects the influence of the traditional culture of akyns (poets). The study revealed key concepts embedded in the phraseological corpus of both languages, as well as the features of their functioning in speech practice. The results obtained can be used in intercultural studies, translation studies and teaching Turkic languages.

Key words: linguistics, phraseological units, national worldview, linguoculturology, Bashkir language, Kazakh language.